Bridging the Gap: Moving to the 1997 Standards for Collecting Data on Race and Ethnicity

Jennifer Madans, Ph.D.
National Center for Health Statistics

14th Annual Morris Hansen Lecture
November 17, 2004

Collaborators

- Bob Anderson
- Elizabeth Arias
- Brady Hamilton
- Deborah Ingram
- David Johnson
- Jacqueline Lucas
- Kristen Miller
- Jennifer Parker
- Nathaniel Schenker
- Paul Sutton
- Stephanie Ventura
- James Weed

Background

- Interest in racial disparities in health

- NCHS data systems are designed to collect data on racial and ethnic populations

- Race specific population data from the Census are used for several purposes at NCHS
  - Denominators for rates
    - Vital statistics
    - Health care provider
  - Survey weights

- Need to obtain consistent comparable race specific population estimates

Differences between the 1977 standards and the 1997 standards

	1977	1997
What are the minimum categories for race?	American Indian or Alaskan Native Asian or Pacific Islander Black White	American Indian or Alaskan Native Asian Black or African American Native Hawaiian or other Pacific Islander White
How many races can be reported?	Only one race	More than one race

Infant mortality rates by race
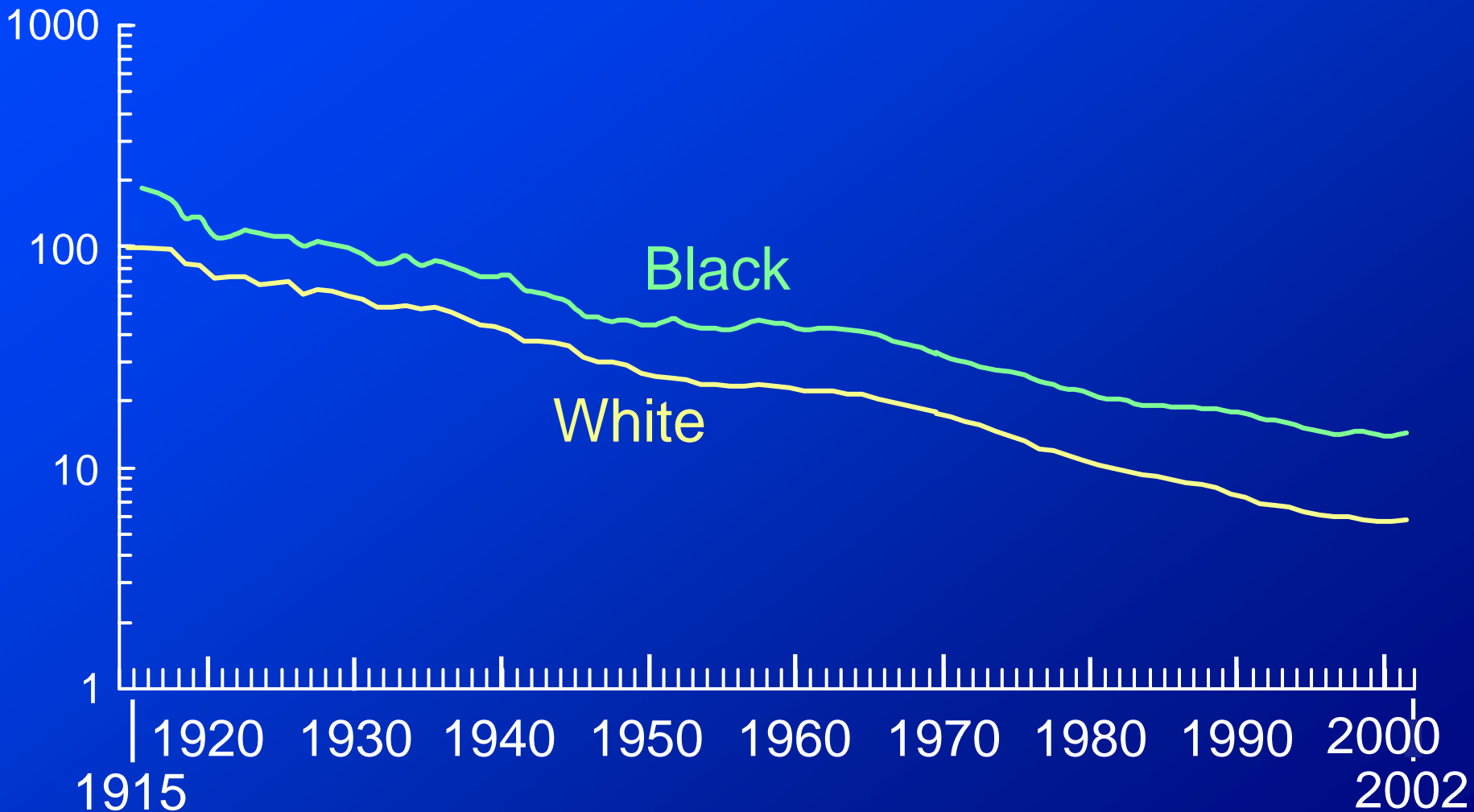
Development of the revised birth & death certificates, 12th revision

- The standard certificates are models developed for use by the individual States.

- The most recent revision took place in 1989.

- In 1998, an expert panel of State vital statistics officials evaluated the current certificates and recommend changes.

- A key recommendation was to incorporate the same questions on race and Hispanic origin that were on the 2000 Census.

Implementation of the Revised Birth and Death Certificates

- NCHS' goal was to have states implement revised certificates soon after the 2000 Census.

- Need to re-engineer their data collection systems to incorporate the latest technology.

- Lack of resources at the State and Federal levels has delayed re-engineering and implementation of the revised certificates.

- Full implementation will be phased in over several years.
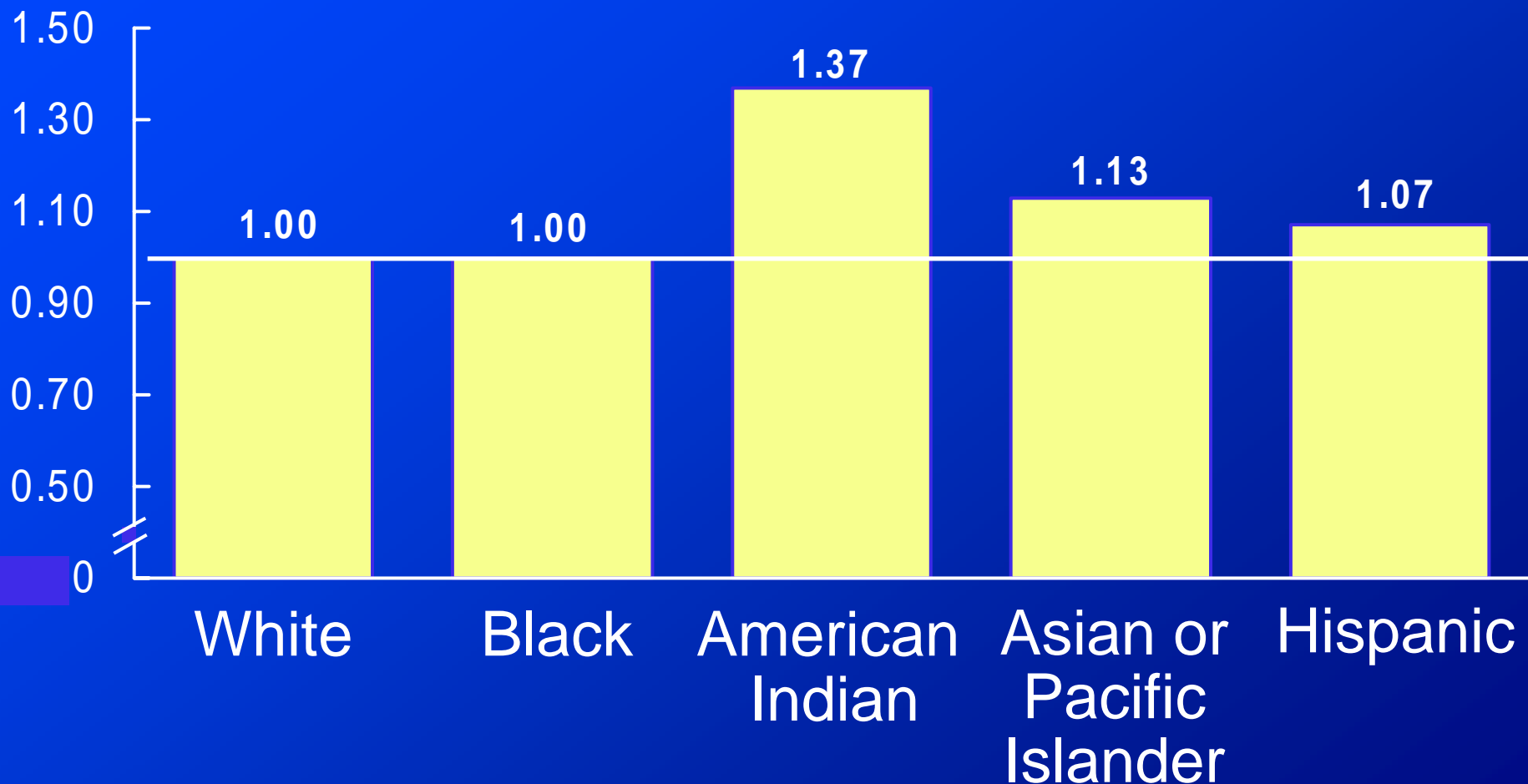
Vital and Health Statistics

September 2001

Series 4, Number 31

Methodological Issues for Vital Rates and Population Estimates: 1997 OMB Standards for Data on Race and Ethnicity

Race misreporting on vital records

Ratio=CPS/DC deaths

	White	Black	American Indian	Asian or Pacific Islander	Hispanic
Ratio	1.00	1.00	1.37	1.13	1.07

NOTE: For Hispanic ratio based on selected States for 1979-85.
SOURCE: CDC/NCHS, National Longitudinal Mortality Study

Bridging between classifications

- Reassigns multiple race responses into a single race category

- Helps explain relationship between old and new data series

- Provides consistent numerators and denominators for transition period

Bridging methods

Criteria for evaluating bridging methods, from Office of Management and Budget (2000)

- Allow correct measurement of change over time
- Be congruent with the respondent's choice
- Have wide range of applicability
- Meet confidentiality and reliability standards
- Minimize disruptions to the single-race distribution
- Be statistically defensible
- Be easy to use
- Require relatively little statistical knowledge
- Be understandable and communicable

Bridging methods (cont'd.)

Bridging methods discussed by Office of Management and Budget (2000), chosen for their simplicity and wide applicability

Deterministic whole assignment

- Largest group
- Smallest group
- Largest group other than white
- Plurality (based on NHIS)

Deterministic fractional assignment

- Equal fractions
- NHIS fractions

National Health Interview Survey (NHIS)

Data source

- Personal interviews

Sample

- 41,000 households annually (approximately 110,000 people)

- Oversample African Americans and Hispanics

National Health Interview Survey (NHIS)

As respondent is handed a card with numbered race groups …

- What is the number of the group or groups which represents---**race?**

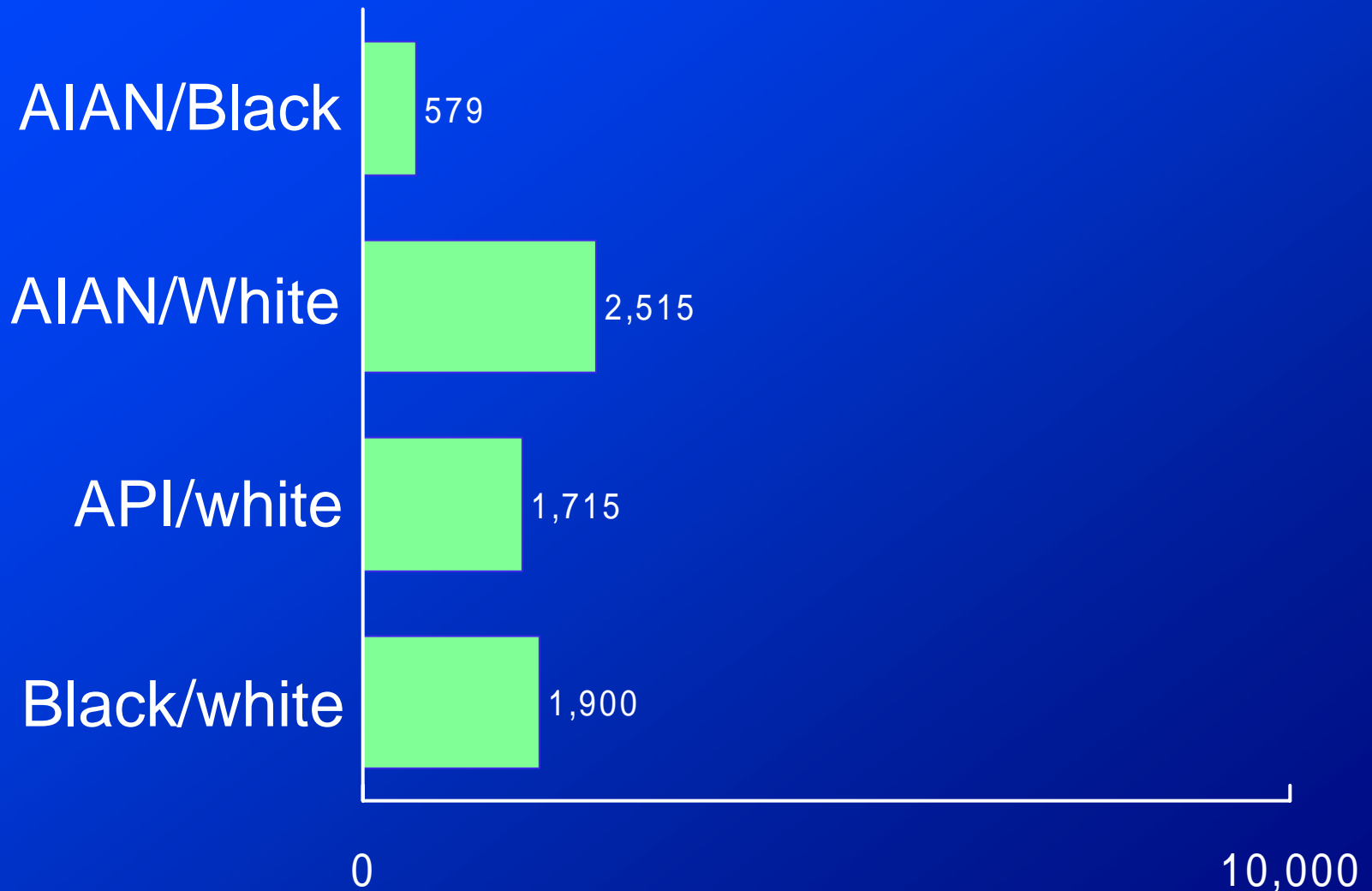- Which of these groups would you say best represents---**race?**

Detailed race: response to initial question

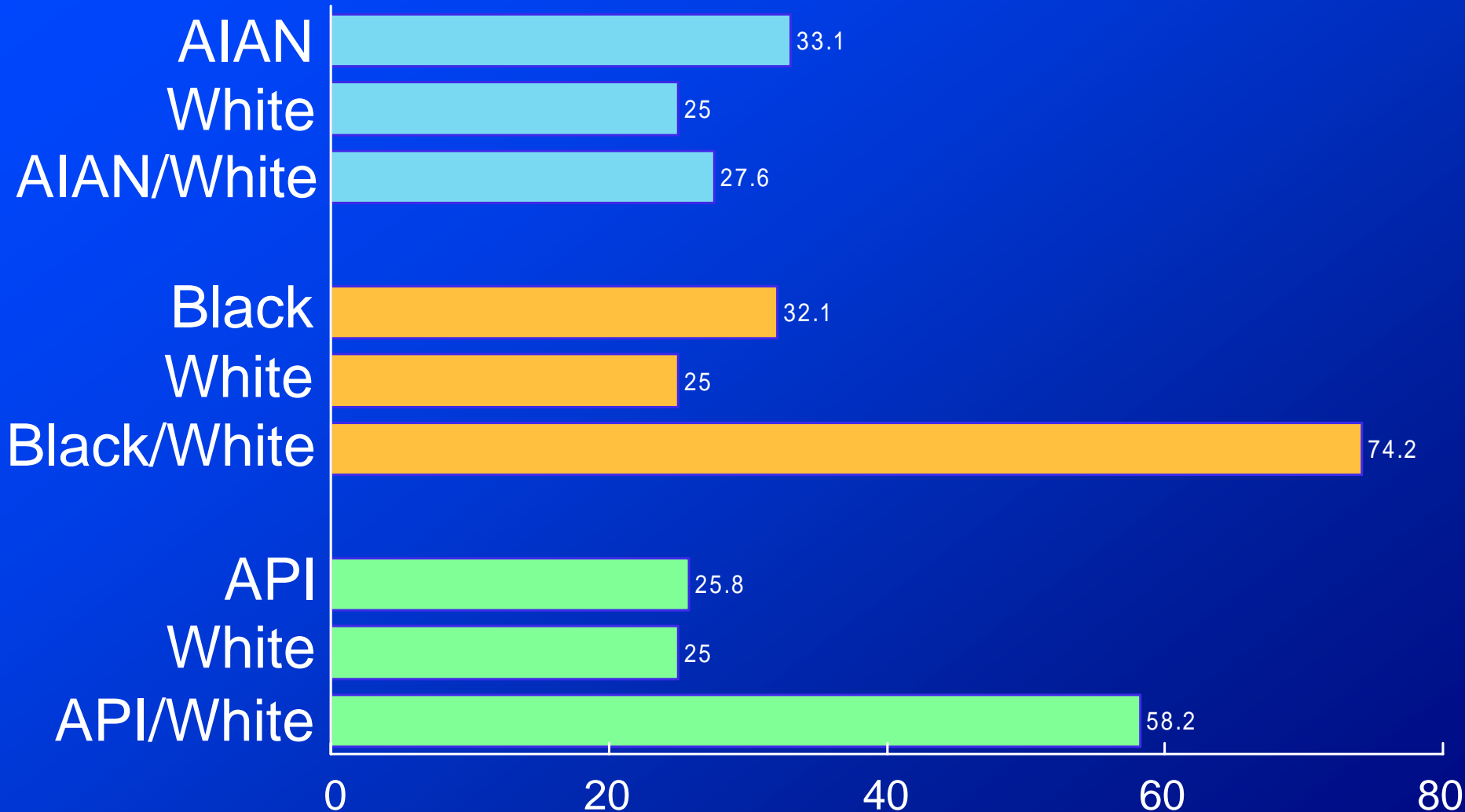**Primary identification**: response to follow-up question

Percent of people reporting more than one race, selected years

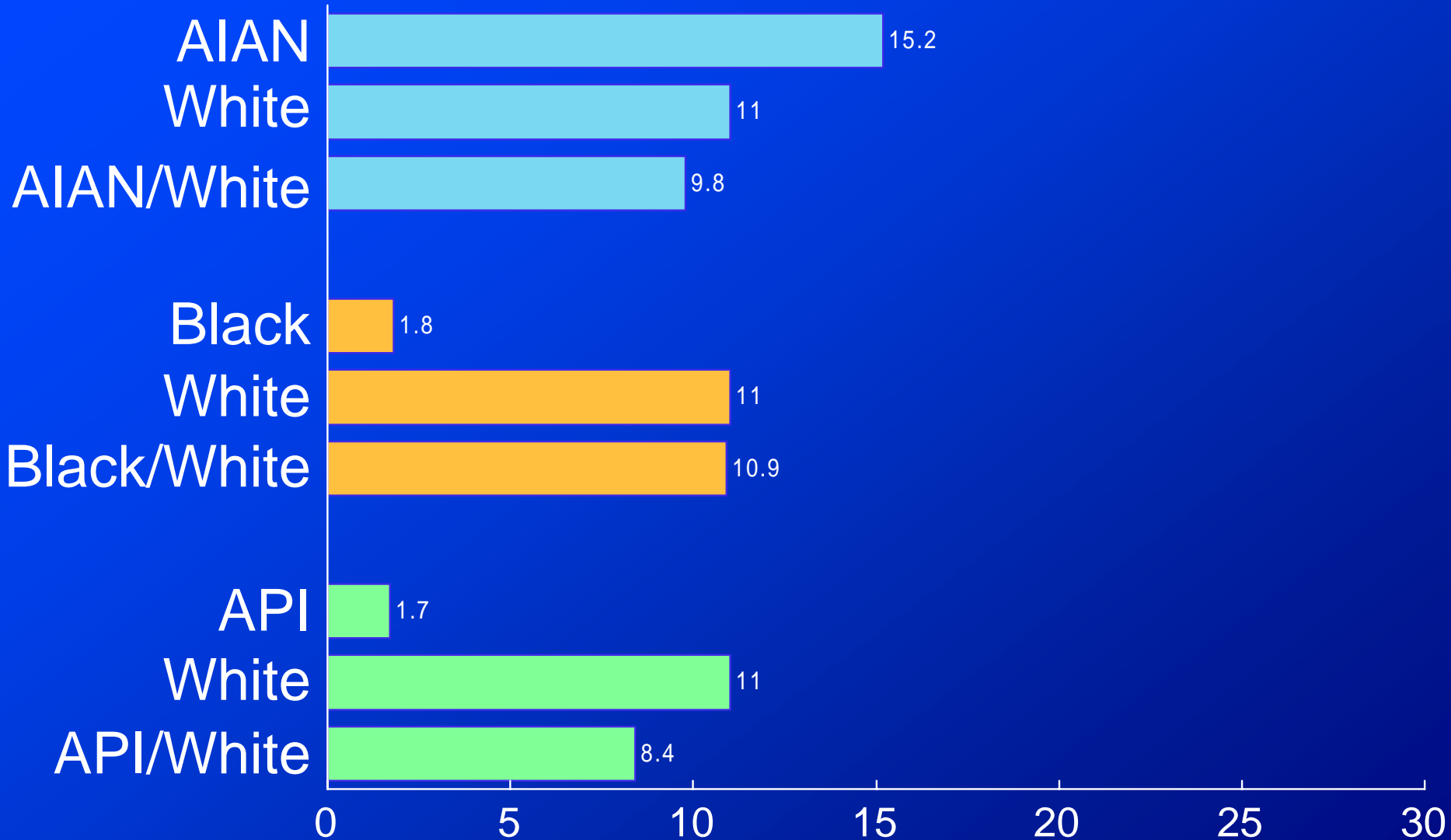| Years	Percent
1982	1.2
1986	1.2
1990	1.3
1994	1.7
1998	1.3
2002	1.6

Number of multiple race respondents: NHIS 1997-2002

Percent under 18 years of age

Category	Value
AIAN	33.1
White	25
AIAN/White	27.6
Black	32.1
White	25
Black/White	74.2
API	25.8
White	25
API/White	58.2

0   20   40   60   80

Percent below poverty

Percent uninsured

Category	Percent
AIAN	36
White	13.3
AIAN/White	23
Black	18.1
White	13.3
Black/White	13.5
API	16.9
White	13.3
API/White	9.9
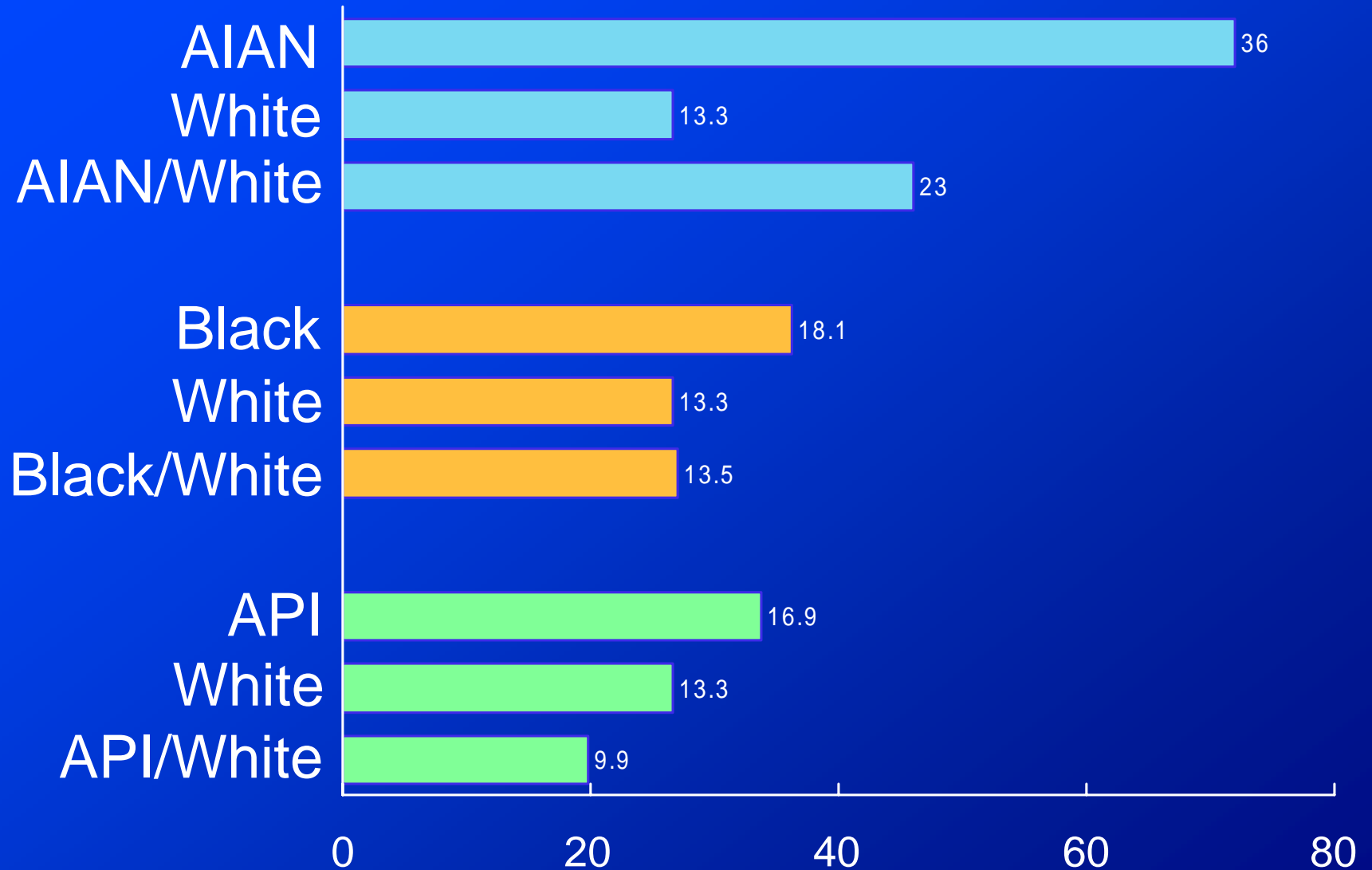
Percent in poor or fair health

AIAN	15.7
White	8.2
AIAN/White	16.1
Black	15.1
White	8.2
Black/White	13.8
API	7.7
White	8.2
API/White	7.8

Age adjusted percent
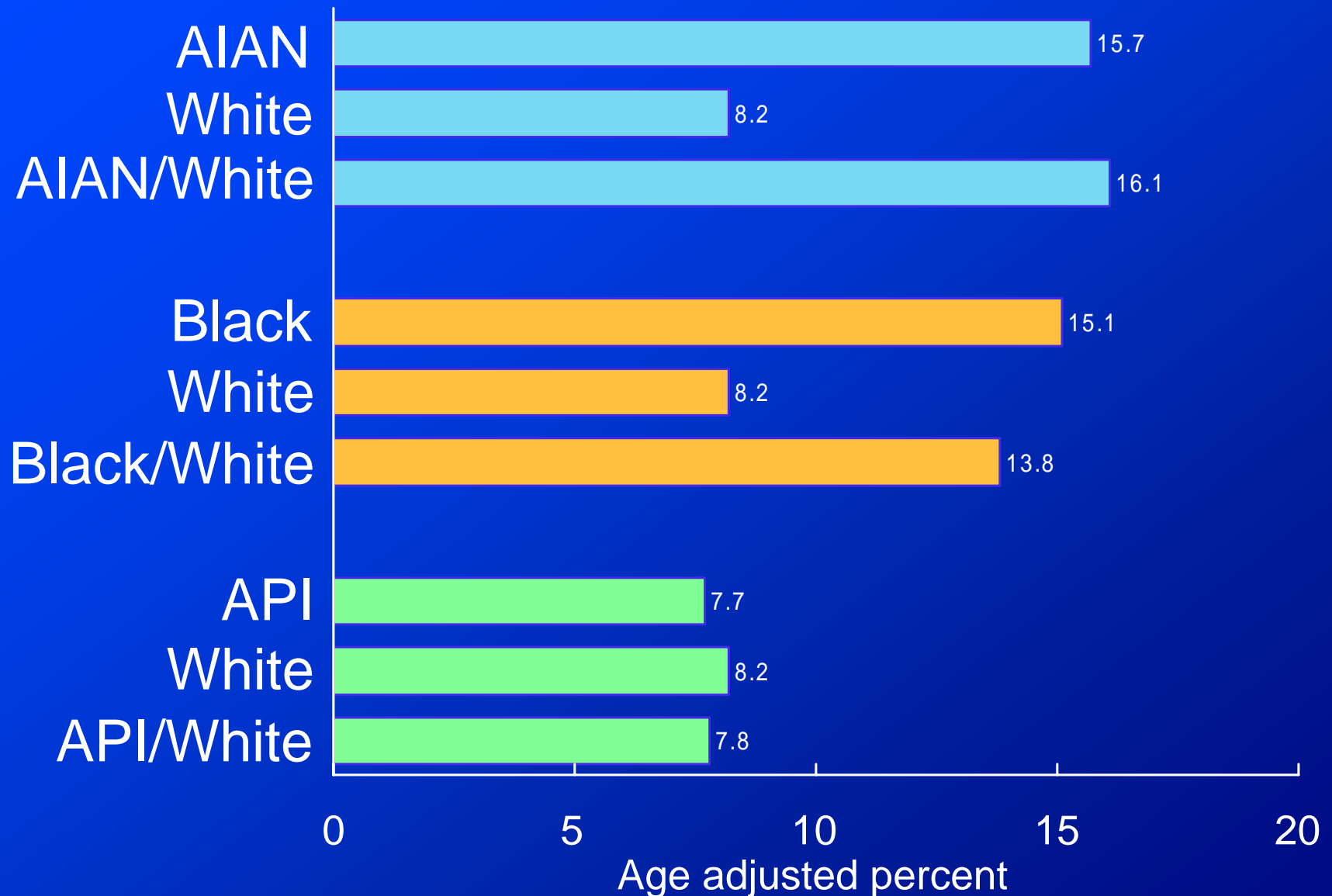
Distribution of primary/identification for multiple race groups
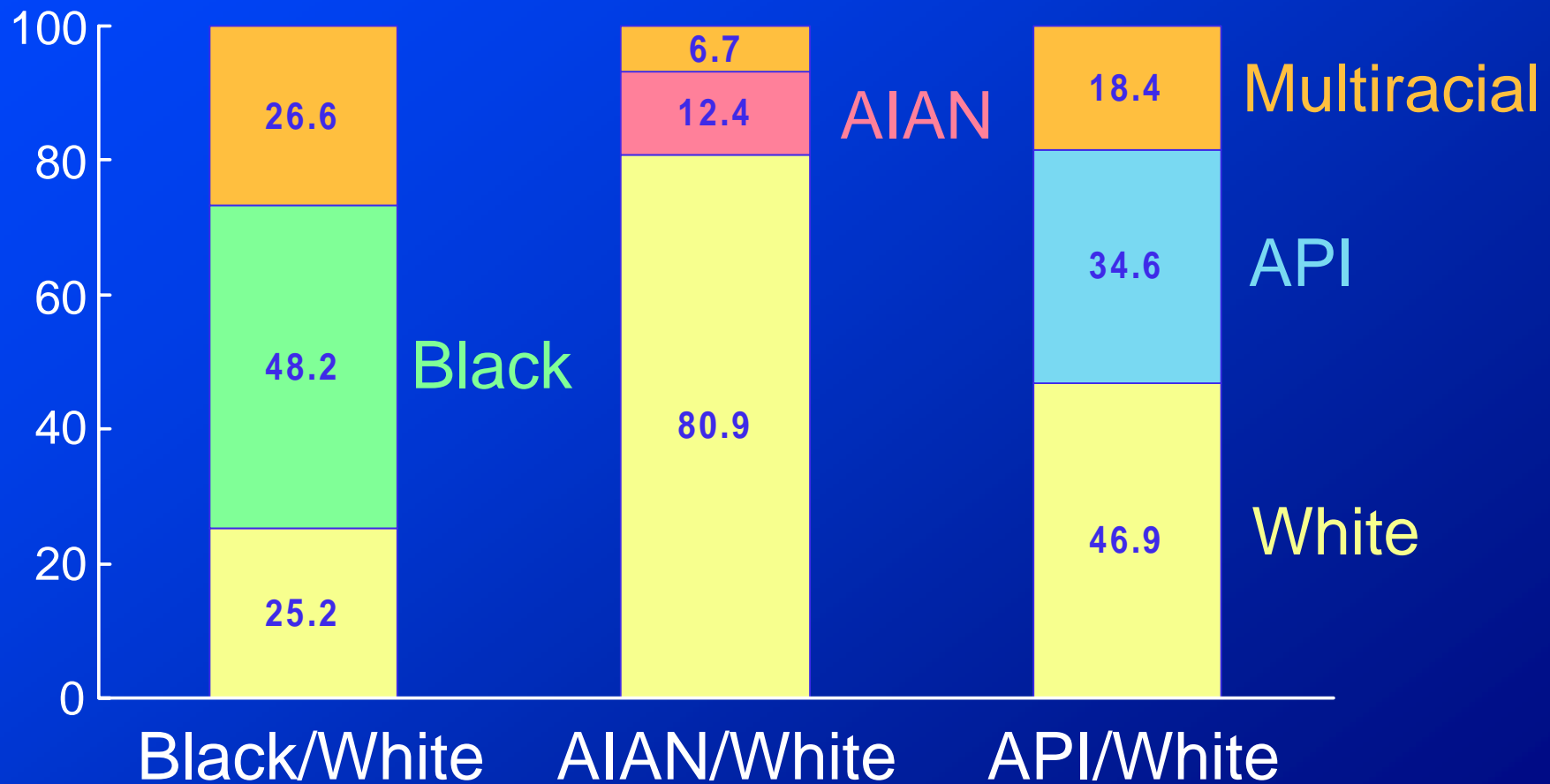
Percent under 18 years of age

Percent Hispanic origin

AIAN — 6.8
White — 10.7
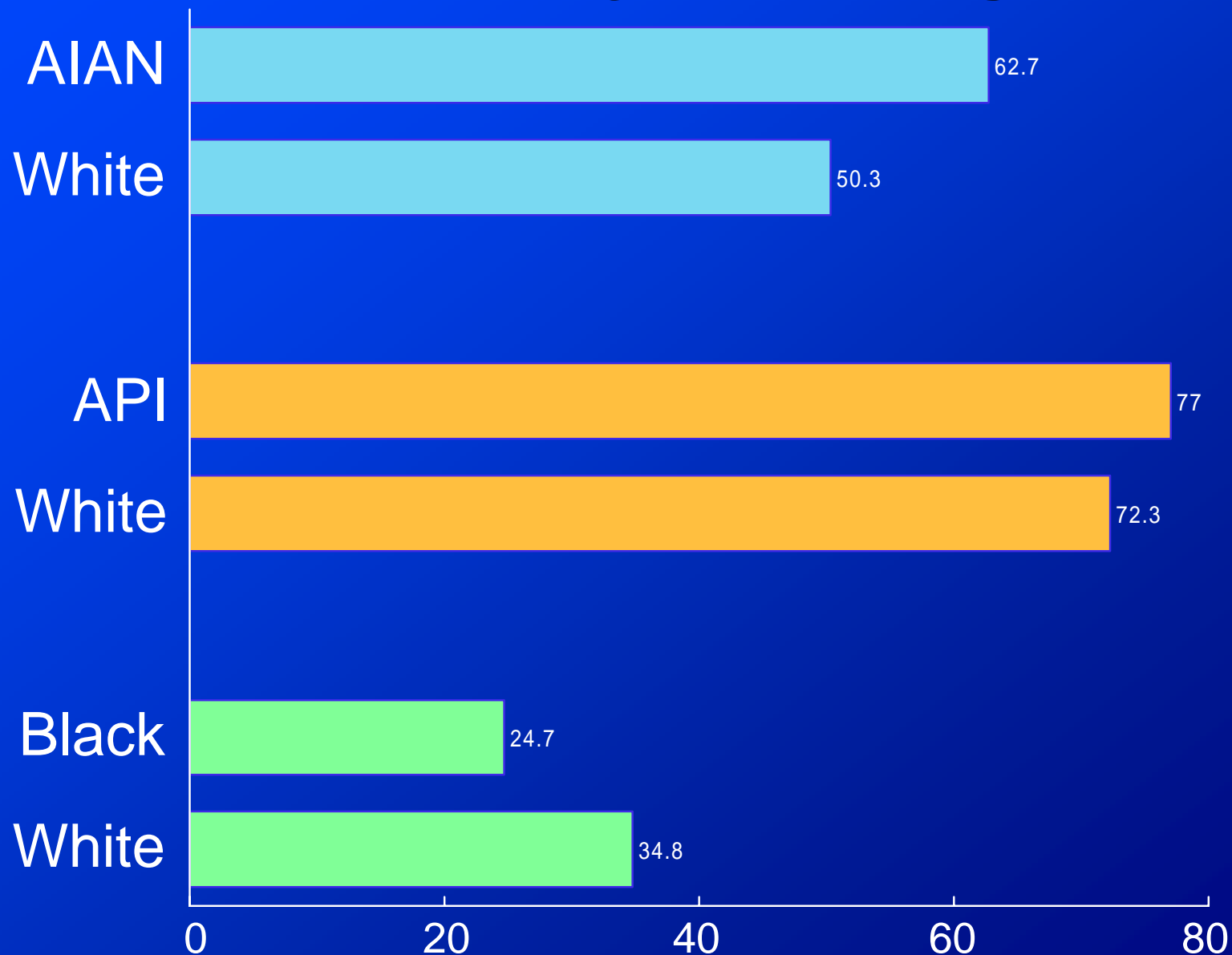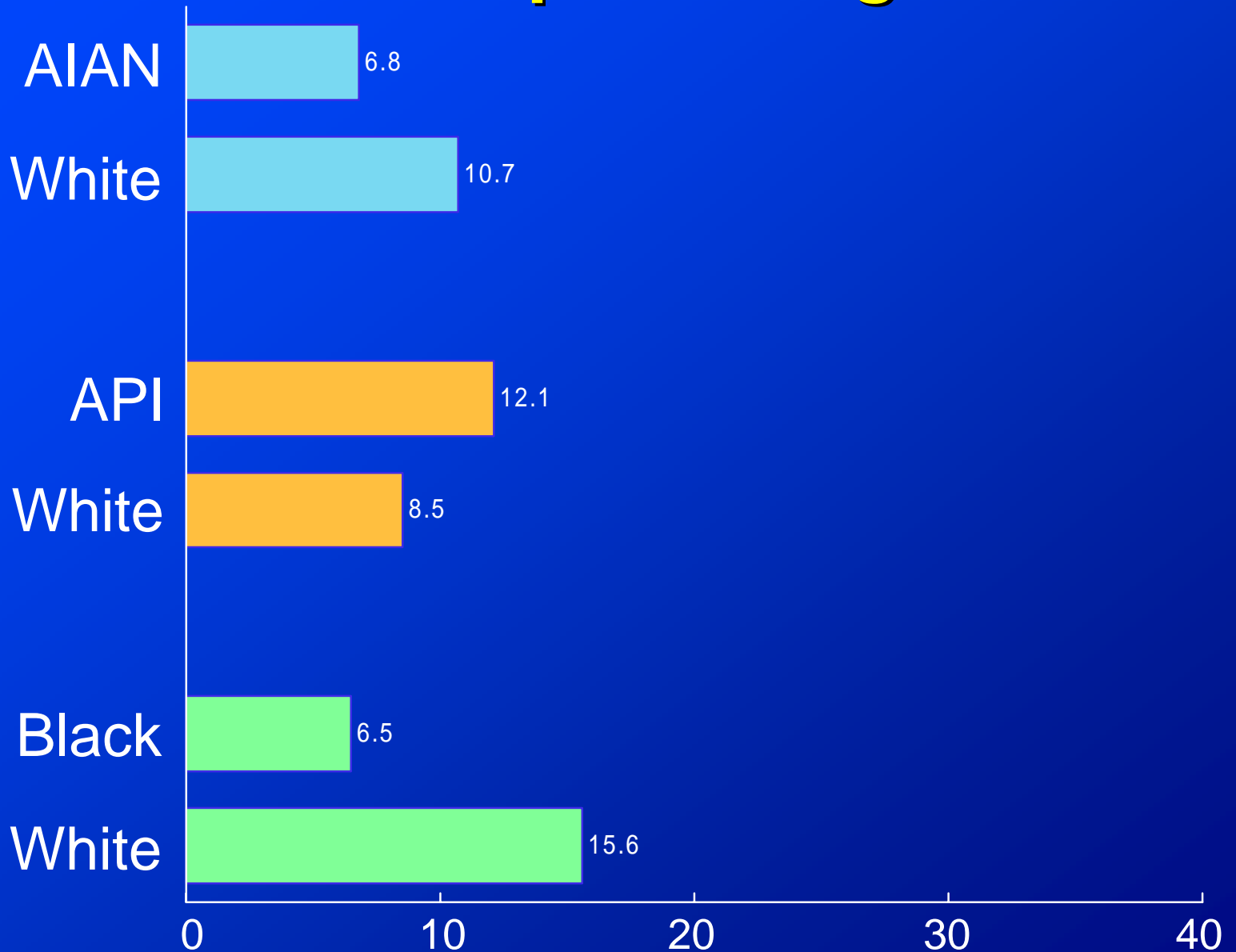API — 12.1
White — 8.5
Black — 6.5
White — 15.6

Percent below poverty

- AIAN — 8.4
- White — 13.8
- API — 29.8
- White — 26.2
- Black — 16.9
- White — 23.8

Percent uninsured

AIAN — 8.4
White — 14.1

API — 12.5
White — 12.5

Black — 20.9
White — 31.7
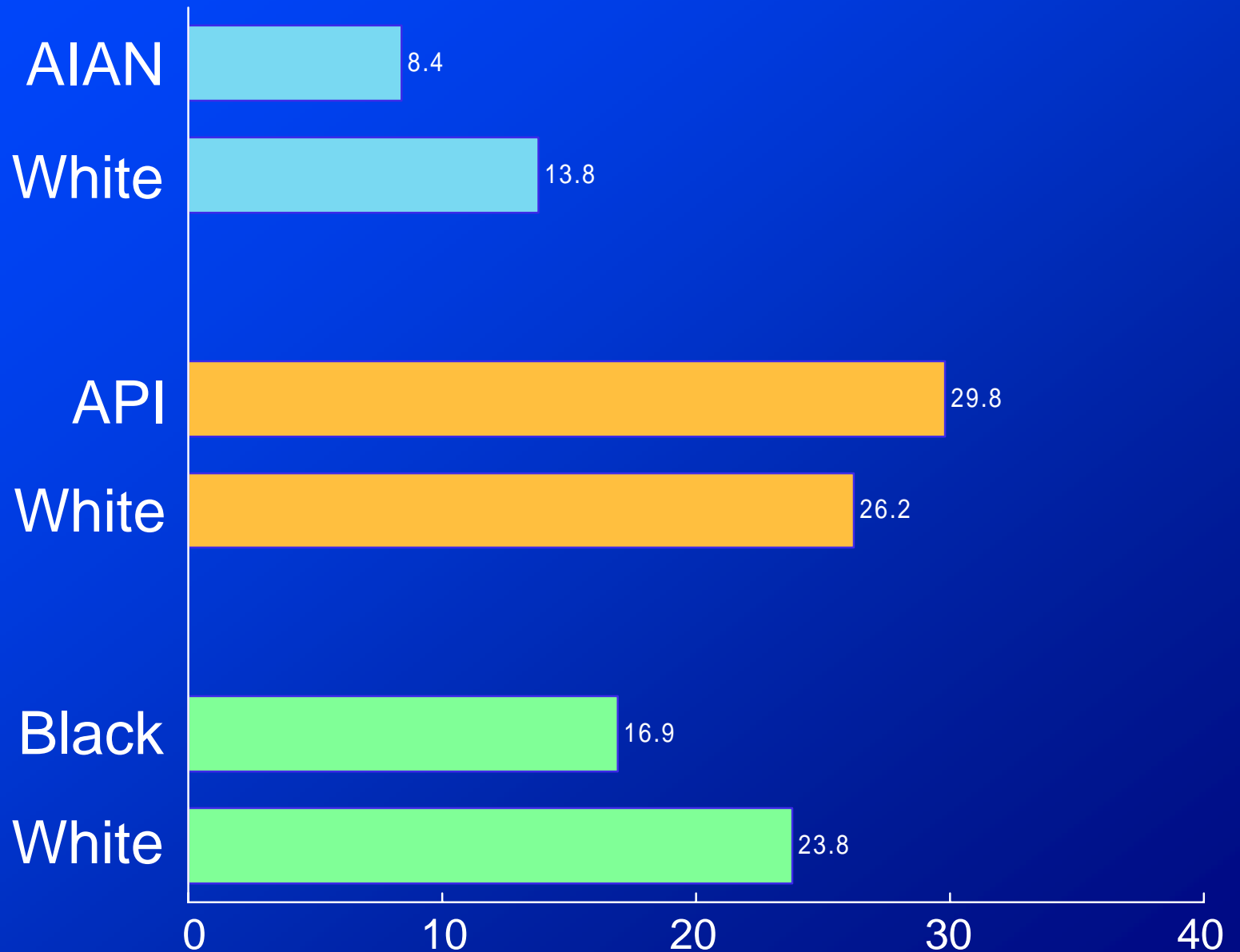
Percent

Percent in fair or poor health status

Category	Age adjusted percent
AIAN	8.6
White	7.2
API	9.8
White	14.8
Black	15.1
White	20.2

Age adjusted percent
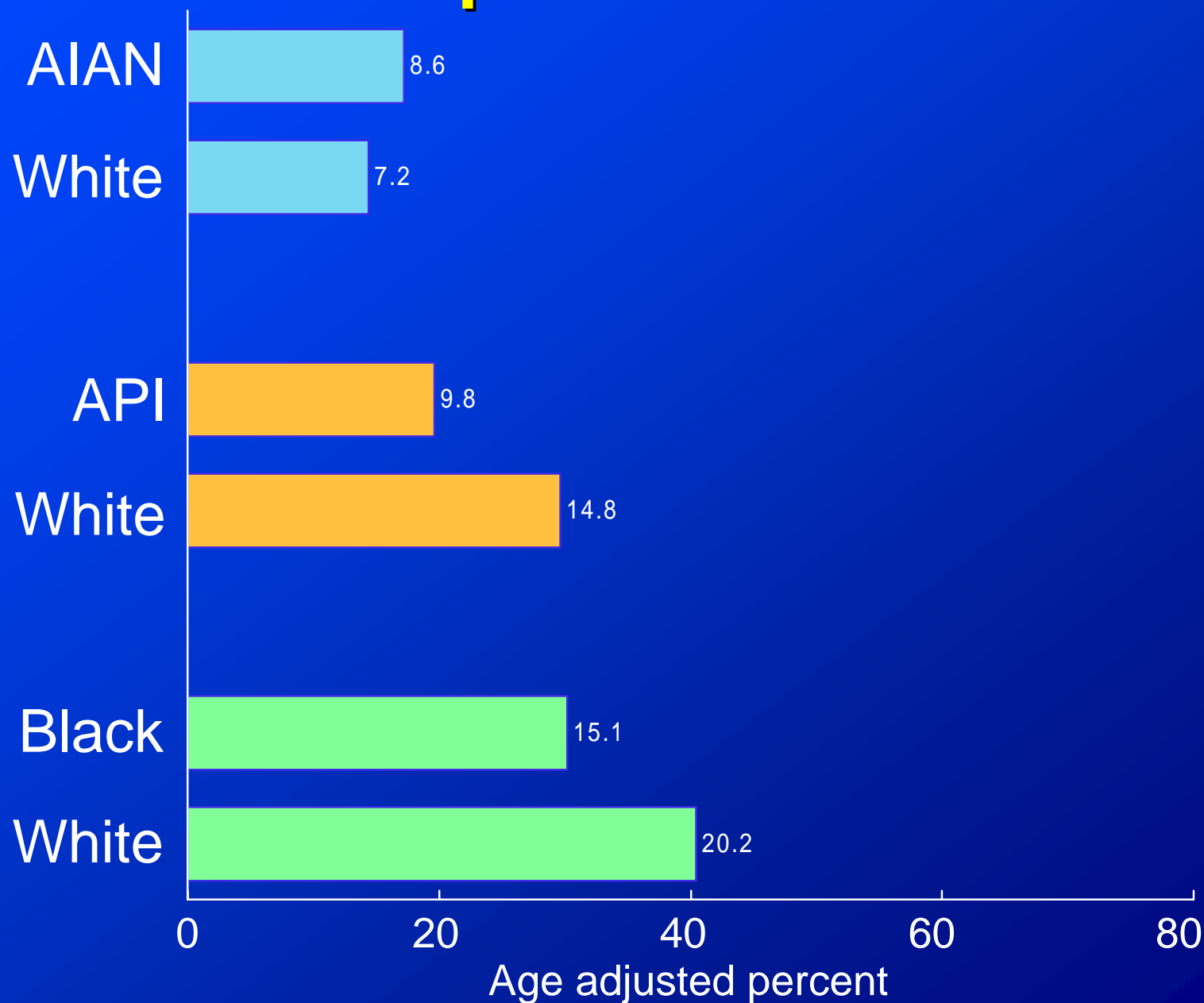
NHIS regression method

- Key assumption: Primary race reported in NHIS follow-up question has a similar distribution (given covariates) to what multiple-race reporters in the census would have reported under the 1977 system

- Categorical regression models fitted to NHIS data from 1997–2000 to predict primary race as a function of personal and county characteristics

NHIS regression method

- Group the multiple-race reporters in the Census 2000 modified race data summary file into multiple-race, county, and person-level covariate combinations.

- Distribute into 1977 race categories in proportion to the estimated probabilities for primary race

- Activities performed by the Bureau of the Census

NHIS regression method

- Predictors included
    - Person-level variables available on the Census 2000 modified race data summary file: age, sex, and Hispanic origin
    - Contextual variables: region, county-specific urbanicity, and race distributions
- Separate models for six largest groups: AIAN and White; API and White; Black and White; AIAN and Black; AIAN, Black, and White; API and Black
- Combined model for remaining groups

Results of modeling

- The model produced estimates that reflected variation across the country
- Models and predictors differed across multiple race groups
- Covariates examined are not very strong predictors of primary race
  - Better predictors probably are not measured in most surveys or the census
- Implications for single race white and black counts are minimal at the National level; AIAN a special case

Bridged and enumerated U.S. population by race: Census 2000

Race	Single race	Total bridged	Percent increase	All inclusive	July 2000, 1990 base
All races	277,668,953	281,421,906	1.4	NA	275,264,999
White	228,104,485	230,085,762	0.9	231,434,388	226,251,833
Black	35,704,124	36,594,309	2.5	37,104,248	35,303,751
AIAN	2,663,818	2,984,150	12.0	4,225,058	2,436,153
API	11,196,526	11,757,685	5.0	12,643,285	11,273,262

Products

NCHS race bridging project, with assistance from the Bureau of the Census

- Created a file of bridged counts for the 2000 census, with multiple-race reporters allocated to the four 1977 race groups
  - Contains population estimates for the four 1977 race groups, within counties, by sex, Hispanic origin, and age
  - Available at:

    www.cdc.gov/nchs/about/major/ dvs/popbridge/popbridge.htm
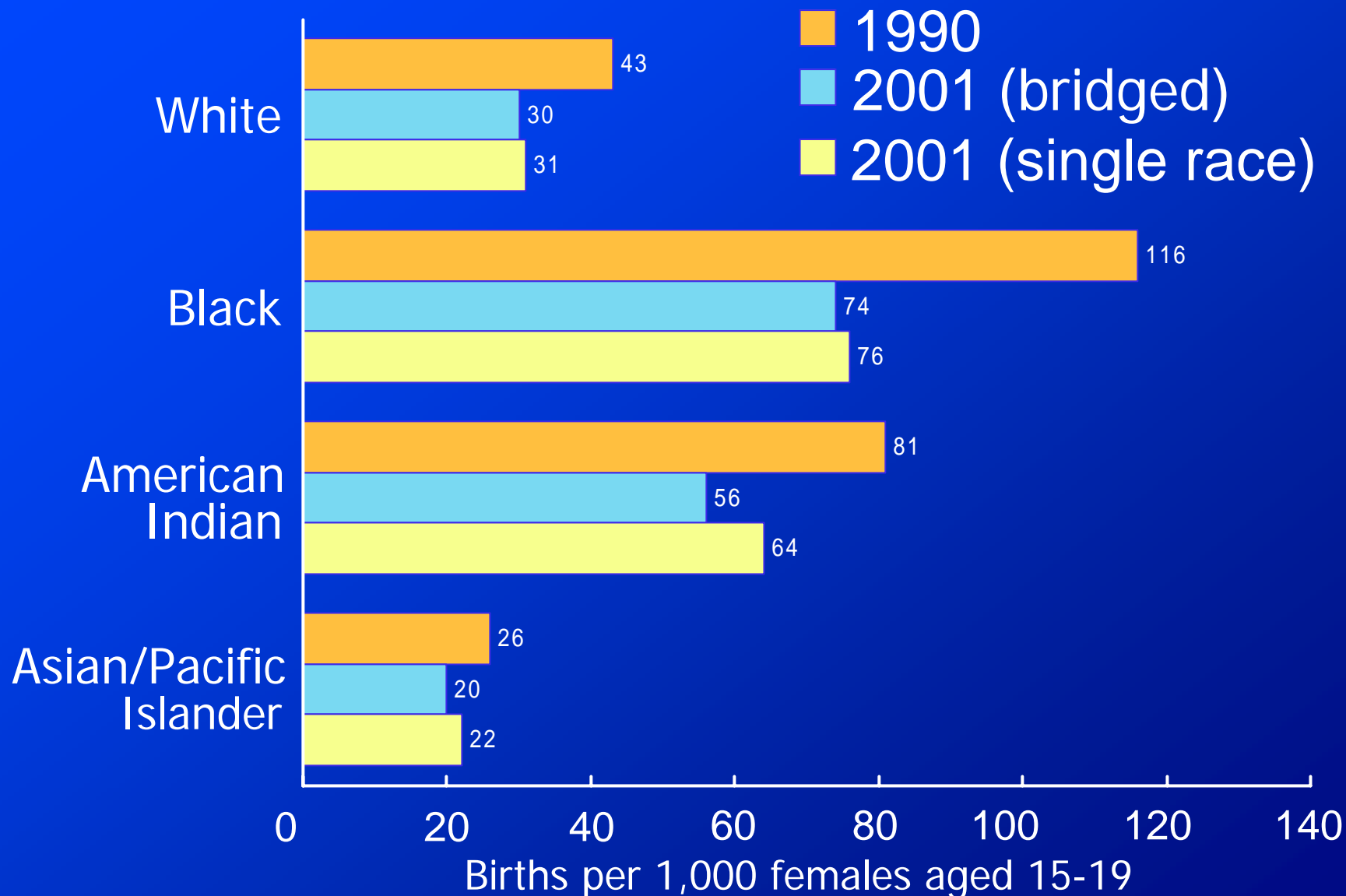
Products (con't)

- Bridged counts also available for intercensal estimates (July 1, 1990 – July 1, 1999) and postcensal estimates (July 1, 2000 – July 1, 2002)
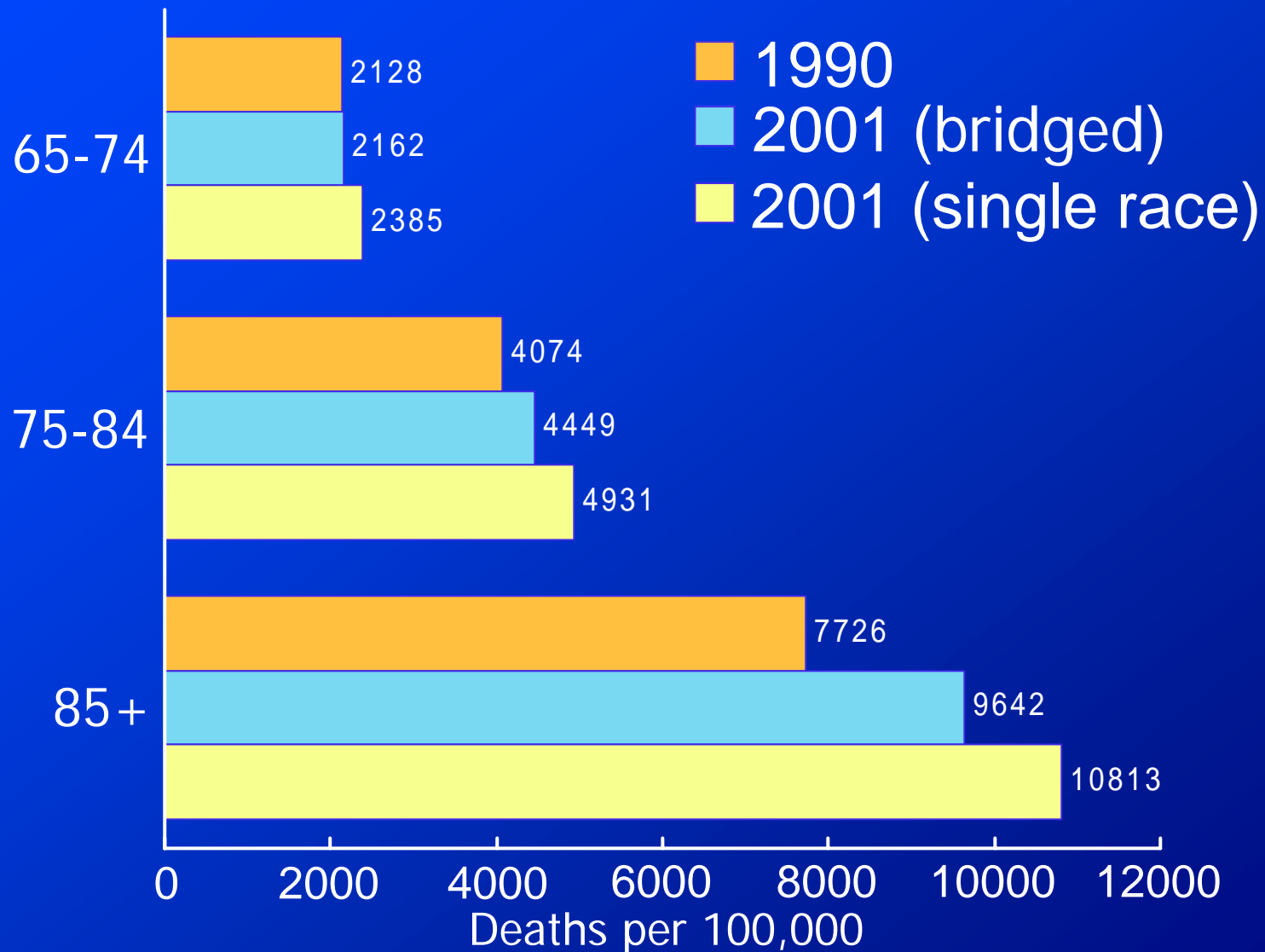
Assessing variability due to bridging

- Although census counts are often considered non-random population quantities, *bridged* census counts have random variability since they are estimates.
- We have developed methods for assessing uncertainty due to bridging
- Bridging does not add substantially to the relative standard errors of birth and death rates at the National level

Using the bridged estimates Teenage birth rates, 1990 and 2001

Legend:
- 1990
- 2001 (bridged)
- 2001 (single race)

White
- 1990: 43
- 2001 (bridged): 30
- 2001 (single race): 31

Black
- 1990: 116
- 2001 (bridged): 74
- 2001 (single race): 76

American Indian
- 1990: 81
- 2001 (bridged): 56
- 2001 (single race): 64

Asian/Pacific Islander
- 1990: 26
- 2001 (bridged): 20
- 2001 (single race): 22

Births per 1,000 females aged 15-19

Using the bridged estimates AIAN death rates

Evaluation: Census Quality Survey

- Use Census Quality Survey to evaluate the bridging models developed at NCHS

- Take advantage of the large sample size of the CQS to explore individual models and alternative approaches for the smaller multiple-race groups

- Improve bridging models by incorporating Census 2000 data such as segregation indexes

Demographic characteristics

Characteristic N	CQS 14,294	NHIS 3,956
	Percent	
Male	49	49
Hispanic	9	9
Age in years		
<18	46	45
18-44	33	35
45-69	18	16
70+	3	3

Census Quality Survey (CQS)

- Some of the CQS and NHIS regression coefficients are similar; others differ in magnitude, direction, and/or statistical significance.

- The resulting NHIS and CQS allocation probabilities are very similar for some multiple-race groups, less similar for others.

- These findings were expected because of the low power of the models.

Evaluation: Do models hold up over time?

NHIS models refit using 1997-2002 data

- **Time covariate**
  - Not significant for black/white or AIAN/white model

Evaluation: Do models hold up over time?

- **Separate models for 1997-2000 and 2001-2002**
  - For AIAN/white and black/white groups, indicators of urban-rural differed between periods
  - For black/white group, the strength of coefficient for percent single race black (squared) increased

- **Models refit with 1) median income; 2)percent population with <high school education; 3) percent foreign born;   4) percent Hispanic—no major differences**

Further implications for vital statistics of the transition

- Full implementation of revised certificates will be phased in over the decade.

- NCHS developed a computer program to code and edit multiple-race data from vital records from both checkboxes and literal entries, by the States.

- This program also bridges the edited race data for parents and decedents to previous 4-race categories using the same bridging algorithm used to bridge population estimates.

Focus on numerators: Multiple-race mortality data for California, 2000-2001
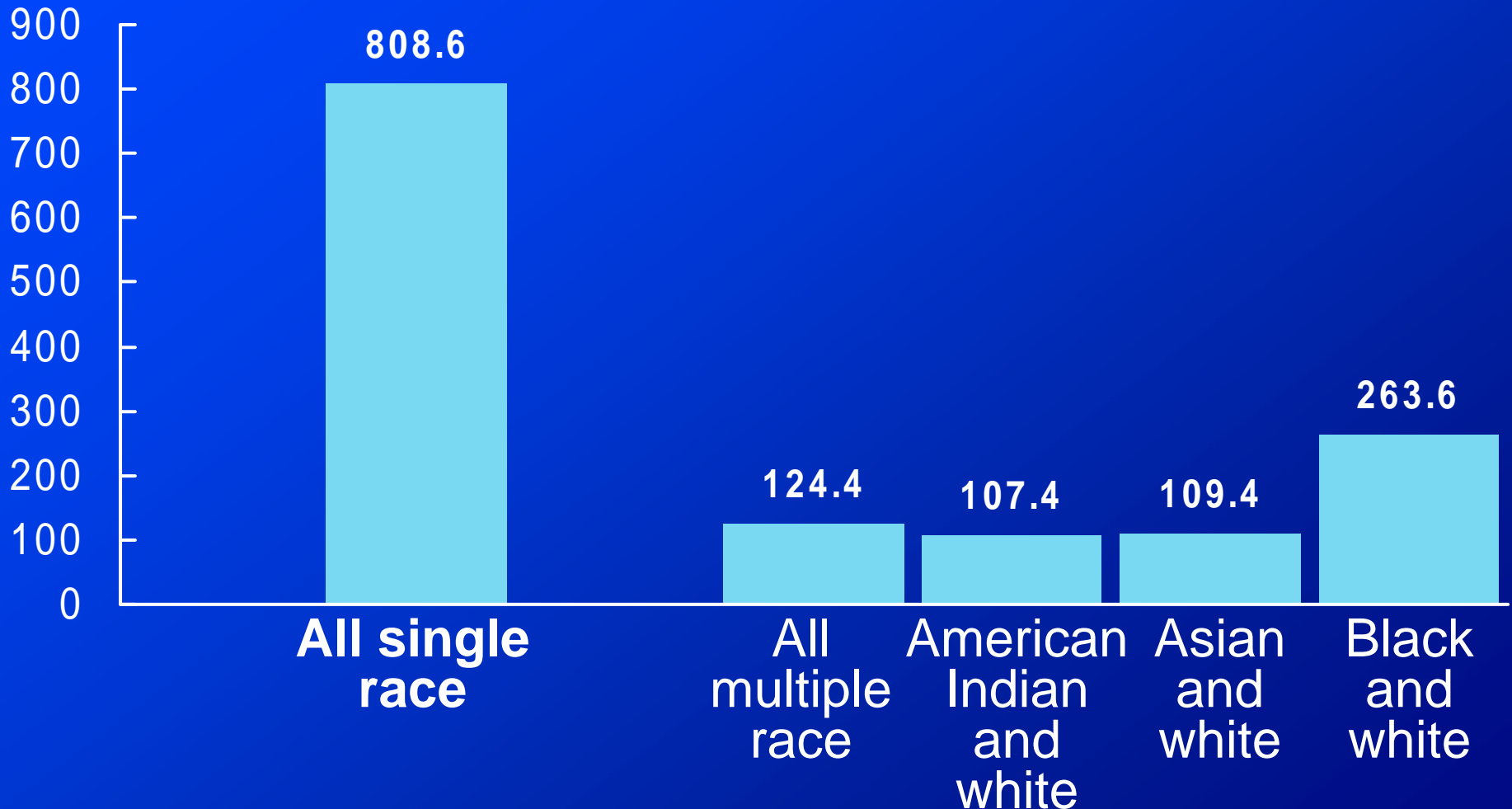
Methods

- Mortality rates for three largest multiple-race groups (African American-white, American Indian/Alaska Native-white, Asian-white) and multiple-race overall using all races recorded on death certificates
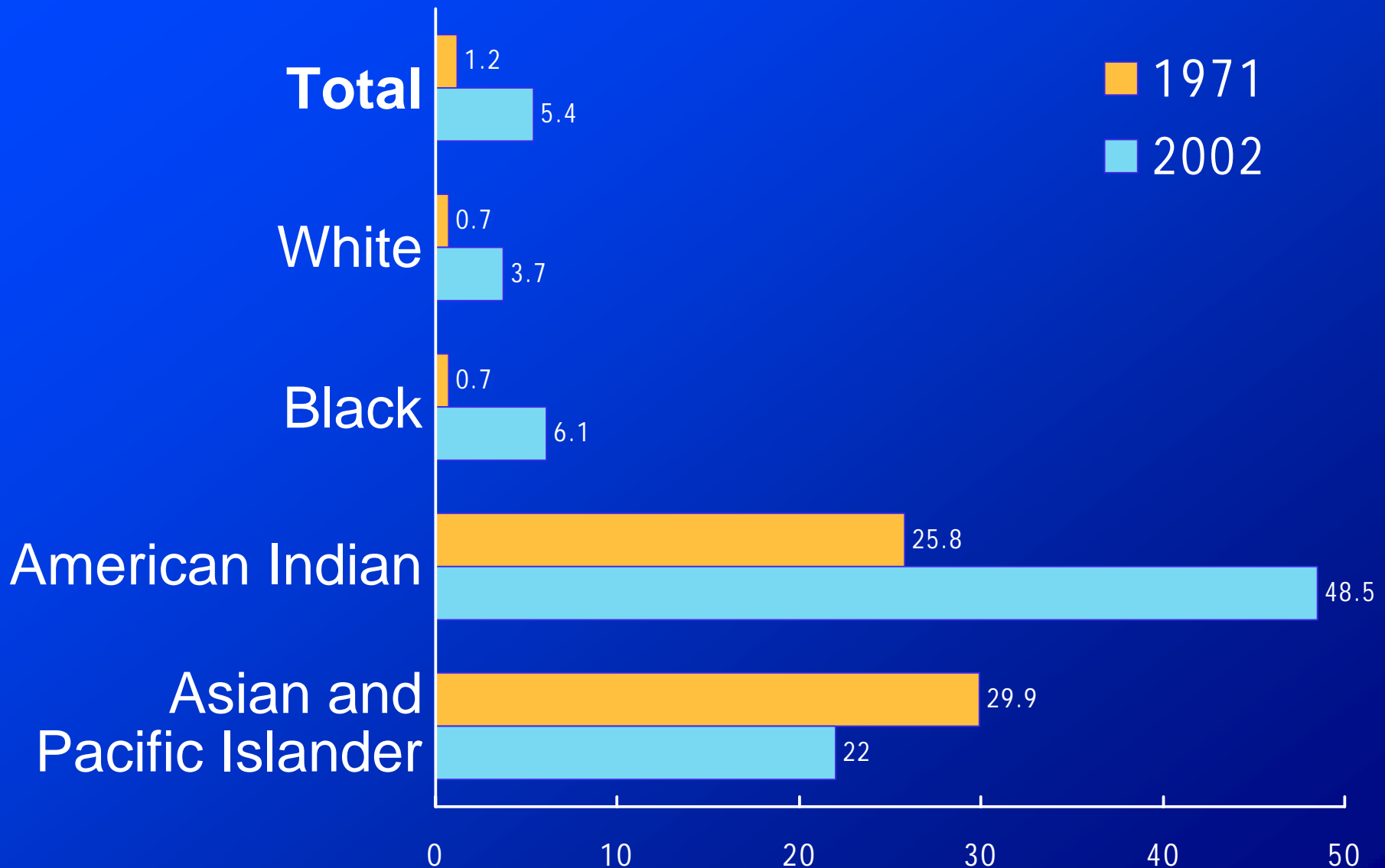
Results

- Multiple-race decedents more likely to be young, Hispanic, male, and never-married
- Age-adjusted mortality rates implausibly low
- Substantial variation by county of residence

All cause age-adjusted death rates by race: California 2000-2001

The future: Interracial births by mother's race

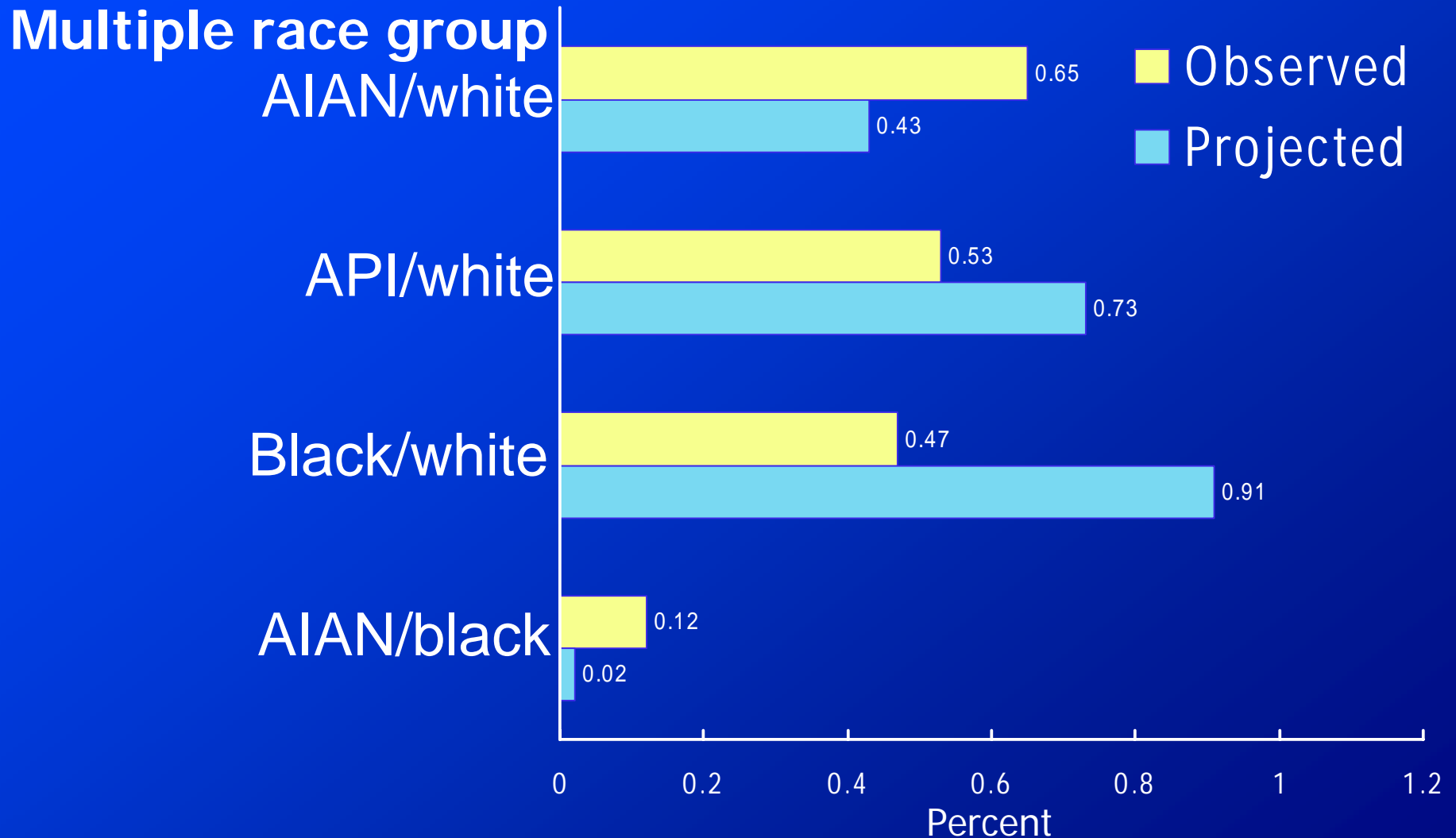# Interracial births and multiple-race reporting

Methods

- Birth certificate data 1968-1998
- NHIS data 1990-1998
- Interracial births compared to multiple-race reporting by race and year of birth.

Results

- Overall multiple-race survey responses corresponded to expectations based on interracial births, but there were discrepancies for specific multiple-race groups
- Generally, fewer black-white survey responses than expected; more AIAN-white and AIAN-black than expected. Differences vary by year of birth.

Overall correspondence between the observed and projected multiple race population
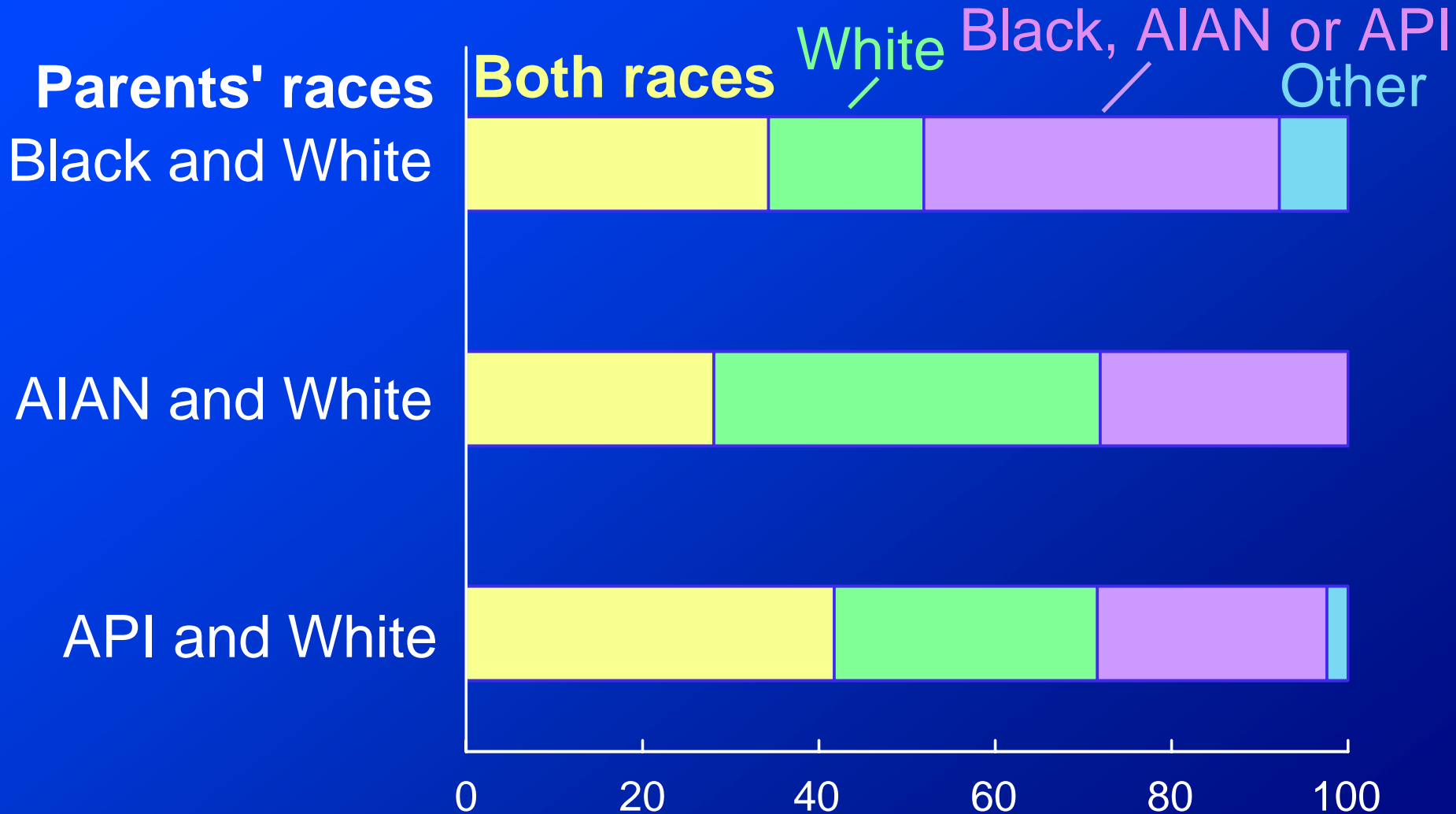
Multiple race group

	Observed	Projected
AIAN/white	0.65	0.43
API/white	0.53	0.73
Black/white	0.47	0.91
AIAN/black	0.12	0.02

Percent (x-axis: 0, 0.2, 0.4, 0.6, 0.8, 1, 1.2)

Percent of children reported to be multiple race by race of father and mother: Washington State, 1999-2002

Race of mother	Race of father				
	Single				
	White	*Black*	*AIAN*	*API*	*Multi-racial*
Single					
White	0.3	82.6	67.0	74.5	81.0
Black	77.0	0.6	75.0	81.7	73.7
AIAN	56.3	84.6	1.0	69.0	70.8
API	75.4	90.1	77.3	0.8	83.3
Multiracial	78.6	81.5	78.4	82.5	95.7

SOURCE: Center for Health Statistics, Washington State.

Reported child's race among children with interracial parents

Parents' races

Both races White Black, AIAN or API Other

Black and White

AIAN and White

API and White

0   20   40   60   80   100

OM133V012

Cognitive testing

- **Social:** based on beliefs about other's perception, usually based on physical characteristics or what others have said

- **Cultural:** based on the community to which there was the strongest sense of belonging

- **Official:** based on the way in which they (or their parents) most often reported their race in administrative or official capacities, such as with birth certificates, driver's licenses, and employment or school applications

- **Ancestral:** based on group composing the largest percentage of their genealogy

Looking back, looking forward

Given what we know should we have built the bridge

- Without a bridge, the 1977 categories would transition to the 1997 single race categories

- The effect would be small for most groups at the National level

- Solution isn't satisfying

Looking back, looking forward

What was achieved

- Better understanding of the change in standards

- A defensible way to transition between standards

- A single set of population estimates available to all users for multiple uses

Looking back, looking forward

What was achieved

- Better understanding of the intercensal period

- Refocusing attention on basic methodologic challenges in collecting data on race and ethnicity

Looking back, looking forward

Where do we go from here

- Need for bridged estimates to calculate vital rates will end

- Misreporting of multiple race on vital and other administrative records presents a harder problem

- Need for better methods for incorporating race into analyses